Supporting Information.

Computational Design, Synthesis, and Evaluation of Miniproteins as Androgen Receptor Cofactor Mimics

Belen Vaz,[a] Sabine Möcklinghoff,[a] Simon Folkertsma,[b] Scott Lusher,[c] Jacob de Vlieg,[b,c] Luc Brunsveld[a,d]

[a] *Chemical Genomics Centre of the Max Planck Society, Otto-Hahn Straße 15, 44227 Dortmund, Germany.*
[b] *Computational Drug Discovery, Centre for Molecular and Biomolecular Informatics, Radboud University, Nijmegen, the Netherlands.*
[c] *Molecular Design & Informatics, Schering-Plough, Oss, the Netherlands.*
[d] *Laboratory of Chemical Biology, Department of Biomedical Engineering, Technische Universiteit Eindhoven, Den Dolech 2, 5612AZ,Eindhoven, the Netherlands. Fax: (+) 31 402478367 E-mail: l.brunsveld@tue.nl*

Contents **Page**

1. General information. S2
2. Folded miniprotein synthesis. S3
3. Synthesis and labeling of the reference peptide. S6
4. Protein Expression and Purification. S8
5. Competitive fluorescence polarization assay. S9

1. General information.

Rink Amide MBHA resin with loading of 0.72 mmol/g and Rink Amide AM resin LL (100-200 mesh) with loading of 0.34 mmol/g were purchased from Novabiochem. Fmoc-protected amino acids were purchased from MultiSyntech and Novabiochem in their appropriately protected form. All other reagents were purchased from Aldrich-Sigma, Fluka and Acros.

All automated peptide syntheses were performed on a Syro II automated peptide synthesizer from MultiSynTech GmbH, using standard *Fmoc*-chemistry. LC-ESI-MS was carried out by using an Agilent 1100 series binary pump together with a reversed-phase HPLC column (Macherey-Nagel) and a Finnigan Thermoquest LCQ. If not otherwise stated, the following gradient program was used for analytical LC-MS: flow: 1 mL/min, solvent A: 0.1% HCO_2H in H_2O, solvent B: 0.1% HCO_2H in CH_3CN, A/B: 90/10 (0-1 min) to 0/100 (over 10 min), 0/100 (12 min). Purification of products by RP-HPLC was performed in an Agilent 1100 Series Purification Platform using a NUCLEODUR® C18 Gravity preparative column from Macherey-Nagel (21 x 250 mm) and flow rate of 25 mL/min. The products were eluted by using different solvent gradients of solvents A and B (solvent A = 0.1% TFA/H_2O; solvent B = 0.1% TFA/CH_3CN). UV signal at 210 nm was used for detection.

2. Folded miniprotein synthesis.

A. Linear miniprotein synthesis (automated).

All sequences were synthesized from *C*- to *N*-terminus on solid support, using an automatic solid phase synthesizer on a 72 μmol scale (100 mg of resin, loading of 0.72 mmol/g). The coupling of amino acids was carried out following standard Fmoc-chemistry, using HOBt/DIC (4 equiv.) as amino acid activation, DMF as solvent and 4 equiv. of the protected Fmoc-amino acids. The resin was first swollen with DMF (1 x 20 min) and the Fmoc protecting group was removed by treatment with piperidine/DMF (2/3, 1 x 3 min; 1/4 1 x 10 min), then washed with DMF (6 x 1 min). One cycle of peptide elongation consisted of the following steps. First, the deprotected resin was treated for 50 min with a cocktail containing the appropriate amino acid (4 equiv, solution 0.3 M in DMF) with equimolar addition of HOBt and DIC (4 equiv, solution 0.3 M in DMF). After washing the resin with DMF (4 x 1 min), the Fmoc protecting group was removed by treatment with piperidine/DMF (2/3, 1 x 3 min; 1/4, 1 x 10 min). After deprotection, the resin was again washed with DMF (6 x 1 min). These steps were repeated until the miniprotein sequence was complete.

After the completion of the sequence, the resin was subsequently washed with DMF (5 x 30 s), CH_2Cl_2 (5 x 30 s) and Et_2O (5 x 30 s) and dried under vacuum for 2-3 h.

Cleavage and side chain deprotection was carried out by treatment of the resin for 3 h with a cleavage cocktail containing TFA/H_2O/EDT/TIS (94:2.5:2.5:1). The cleaved resin was washed with TFA (3 x 15 s) and the cleaved linear miniprotein was collected, concentrated by rotary co-evaporation with toluene into less than 1 mL solution and

precipitated by addition of cold Et_2O (10 mL). The mixture was cooled in a liquid N_2 bath for 1 min, centrifuged (4000 rpm, 5 min, 4 °C) and the Et_2O was decanted from the pellet. Cold Et_2O was added again and the procedure was repeated twice. The crude peptide obtained was dissolved in H_2O/CH_3CN and lyophilized to dryness.

B. Folding and purification of miniproteins.

The crude products obtained from lyophilization were dissolved in a mixture 2:1 of sodium phosphate buffer (100 mM NaH_2PO_4/Na_2HPO_4, 1:17.8, pH 8) and trifluoroethanol (TFE) to a final concentration of 0.5 mg/mL. In order to reduce previously formed random disulfide bridges, the solution was stirred with 2 equiv. of TCEP for 2 h at room temperature. After addition of 10 vol% of dimethylsulfoxide (DMSO), subsequent oxidation was carried out by exposure to air for 24-48 h. The transformation was followed by LC-MS. When the oxidation was complete, the TFE was removed by evaporation and the resultant aqueous solution was lyophilized.

The crude product was dissolved in water and the salts were removed by filtration through a SepVak® Vac C18 column. After elution of the salts with 2 volumes of water, the oxidized miniprotein was eluted with a mixture of H_2O/CH_3CH (1:1) and then pure CH_3CN. Fractions containing acetonitrile were combined and lyophilized.

The crude oxidized miniproteins were purified by reverse-phase HPLC on a Nucleodur C18 Gravity column (125 x 21 mm, Macherey-Nagel) with a linear gradient of A (0.1% TFA in H_2O) and B (0.1% TFA in CH_3CN) from 10% of A to 40%-60% of B and flow rate of 25 mL min^{-1} and were detected at 210 nm using a diode array UV/VIS detector. The identities and purities of the purified folded miniproteins were assessed by LC-MS (ESI mass spectrometry) (Table S1). The folding of the miniprotein was assessed by comparison of their CD spectra with those of the natural miniprotein. Following purification, all miniproteins were lyophilized and kept at –20 °C.

Table S1. Analytical data of synthesized miniproteins.

Miniproteins	Mass calcd $(M+2H)^+$	Masses found $(M+2H)^+$	Purity	Yield
Apa-0	1013.96	1014.38	94%	12.5%
Apa-1	1004.92	1005.26	95%	2.8%
Scy-0	1366.40	1366.00	> 90%	3.6%
Scy-1	1391.11	1391.76	> 90%	0.5%
Hef-0	1327.54	1327.83	> 90%	7.4%
Hef-1	1282.00	1282.42	> 95%	12.8%
Hef-2	1184.97	1185.34	> 99%	11.3%
Hef-3	1351.58	1352.20	> 90%	1.8%
Omt-0	1258.51	1258.88	> 99%	3.0%
Omt-1	1280.06	1280.41	85%	0.8%
Omt-2	1127.49	1127.74	85%	2.6%
Omt-3	1269.03	1269.53	85 %	1.0%
Reference (Fl)	1110.10	1110.53	> 99%	3.5%

3. Synthesis and labelling of the reference peptide.

A high affinity peptide for the androgen receptor was obtained by Fletterick *et al*[1] by phage display (One of the best known short peptide AR binders). This peptide was selected for the competitive fluorescence polarization assay as the competitive CBI for the LBD-AR. An analogue of this peptide, featuring an *N*-terminal cysteine, was synthesized, allowing the introduction of a fluorescein label. The synthesis of the unlabeled peptide was carried out on solid support using an automatic solid phase synthesizer and following the method described in page S3 for the synthesis of the linear miniproteins.

SSRFESLFAGEKESR

Fluorescein-CSSRFESLFAGEKESR

The peptide featuring a cysteine in the *N*-terminus was labeled with fluorescein. For this, a solution of the peptide in potassium phosphate buffer (100 mM, pH 7.2, previously degassed by sonication for 1h at r.t., 1mg/mL) was prepared and TCEP·HCl (10 equiv.) was added. The resulting clear solution was stirred for 1h at r.t. After this time, a solution of 4(5)-(Iodoacetamido)fluorescein (5 equiv.) in DMSO (10 mg/mL) was added to the reaction mixture and the homogenous mixture was stirred at room temperature. The course of the reaction was followed by LC-MS. After 42 h no unlabeled peptide was detected, and the reaction was quenched with ethanethiol. The resulting mixture was directly lyophilized.

The labeled peptide was purified by RP-HPLC (NUCLEODUR® C18 Gravity column, Macherey-Nagel). The eluents used were A: H_2O (+0.1% HCOOH) and B: CH_3CN (+0.1% HCOOH). The method featured a 20 min. gradient of a mixture

A/B 90:10 to A/B 50:50 affording the expected fluorescein labeled peptide with high purity (>99 %) and 10% yield for the labeling process.

The concentration of the solution of the Fl-peptide in HEPES buffer was determined by UV absorption at 495 nm (molar absorption of fluorescein 75000 M^{-1} cm^{-1}). This concentration was determined by four different dilutions of the original Fl-peptide solution, measuring the absorption value at 495 nm (the zero values were corrected by the corresponding values measured at 580 nm).

4. Protein Expression and Purification.

The hAR-LBD (residues 664-919) was subcloned into pGEX-KG with an N-terminal GST tag and was a kind gift from P. Donner (Bayer-Schering Pharma AG). The plasmid was transformed into *Escherichia coli* BL21 (DE3). After being induced with 30 μM IPTG at OD_{600} = 1.2, the cells were grown in the presence of 10 μM DHT (5α-Androstan-17β-ol-3-one, Fluka) for 18-20 h at 18 °C and harvested by centrifugation at 4500 rpm, 20 min. The cells were lysed with a microfluidizer (4 passes at 600 kPa) in buffer H (50 mM HEPES, pH 7,3, 300 mM NaCl, 5 mM EDTA, 10% glycerol, 100 μM DHT, 100 μM PMSF, 10 mM DTT) and centrifuged at 20.000 rpm for 30 min. The soluble cell lysate was immobilized on a glutathione Sepharose 4 Fast Flow affinity matrix (Amersham Biosciences), washed with buffer A (50 mM HEPES, pH 7,3, 300 mM NaCl, 5 mM EDTA, 10% glycerol, 10 μM DHT, 1 mM DTT) and eluted with buffer A containing 15 mM glutathione. Fractions containing the fusion protein were combined and desalted on a Sephadex G25 PD-10 column (Amersham Biosciences) preequilibrated with buffer A. In case of cleavage of the GST moiety with thrombin over night at 4 °C a final purification step was followed using a HiTrap SP cation exchange column (Amersham Biosciences). The concentration of the protein was quantified using the Bio-Rad protein assay kit with β-Tubulin as the standard.

5. Competitive fluorescence polarization assay

-Setup of fluorescence polarization assay

Purified hAR-LBD and fluorescein-labeled peptide Fl-CSSR**F**ES**LF**AGEKESR (Fletterick *et al*. *PLoS Biology*, **2004)** were used for the fluorescence anisotropy assay. A 50 µl stock solution of fl-peptide (0,1 µM) in Assay-Buffer (20 mM HEPES, pH 7,2, 50 mM KCl, 1 mM EDTA, 10 µM DHT, 1 mM DTT, 1mg/ml BSA) was placed in separate wells of a black 384-well plate (Perkin Elmer, Optiplate-384 F) containing increasing concentrations of hAR-LBD. Wells containing only stock solution without protein were used as negative control. The mixture was incubated at room temperature for 1 h in the dark. The fluorescence anisotropy was measured at 23 °C using a plate reader (SpectraMax M5, Molecular Device) with a 485/494 nm excitation and a 535/525 nm emission filter. The polarization (in millipolarization, mP) was plotted against increasing concentrations of the hAR-LBD and the resulting sigmoid curve was then fitted with a Hill binding model using ORIGIN 7 (Scientific Graphing and Analysis Software, OriginLab Corp.) to determine the K_D value of the fluoresceine-labeled peptide - hAR-LBD complex (Figure S1). With these experimental data and the use of Equation 1 a Kd = 0.89 µM was determined for the fluorescent reference peptide.

$$P = P_0 + \frac{(P_{fin} - P_0)\left(A_0 + B_0 + K_d - \sqrt{(A_0 + B_0 + K_d)^2 - 4A_0B_0}\right)}{2A_0} \quad \text{(Eq. 1)}$$

P_{fin} and P_0 : maximum and minimum values observed for the fluorescence polarization; A_0 : total concentration of labeled peptide; B_0 : total concentration of protein; K_d : dissociation constant of the protein-peptide complex

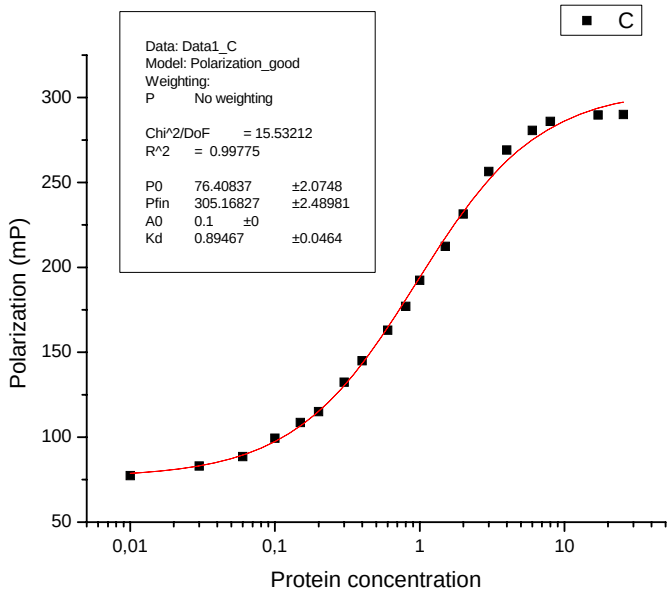

Figure S1. Polaritazion curve of the titration of the AR-LBD to the fluorescein labeled reference peptide. Polarization values in mP are plotted against the $\log_{10}$ of the protein concentration. The resulting sigmoid was fitted to the equation 1.

-Competitive fluorescence polarization assay

The optimized reaction mixture contained 0,1 µM fluorescein-labeled peptide Fl-CSSR**F**ES**LF**AGEKESR (Fletterick *et al*. *PLoS Biology*, **2004)** and 1 µM of purified GST-hAR-LBD in Assay-Buffer (20 mM HEPES, pH 7,2, 50 mM KCl, 1 mM EDTA, 10 µM DHT, 1 mM DTT, 1mg/ml BSA). Inhibition experiments were performed in 384-well plates (Perkin Elmer, Optiplate-384 F) by adding 40 µl of the reaction mixture to 10 µl of increasing amounts of inhibitor (diluted in Assay-Buffer). Reaction mixtures without inhibitor as well as mixtures containing only Fl-labeled peptide in Assay buffer were used as controls. The mixture was incubated at room temperature for 1 h in the dark. The fluorescence anisotropy was measured at 23 °C using a plate reader (SpectraMax M5, Molecular Device) with a 485/494 nm excitation and a 535/525 nm

emission filter. The polarization (in millipolarization, mP) was plotted against increasing concentrations of the miniprotein ($2 \cdot 10^{-9}$ M to $2 \cdot 10^{-3}$ M) and then fitted with a Klotz binding model (Klotz *et al.* 1946) to a sigmoid curve using ORIGIN 7 (Scientific Graphing and Analysis Software, OriginLab Corp.) to determine the and IC_{50} value K_i of the inhibitor (Figure S2).

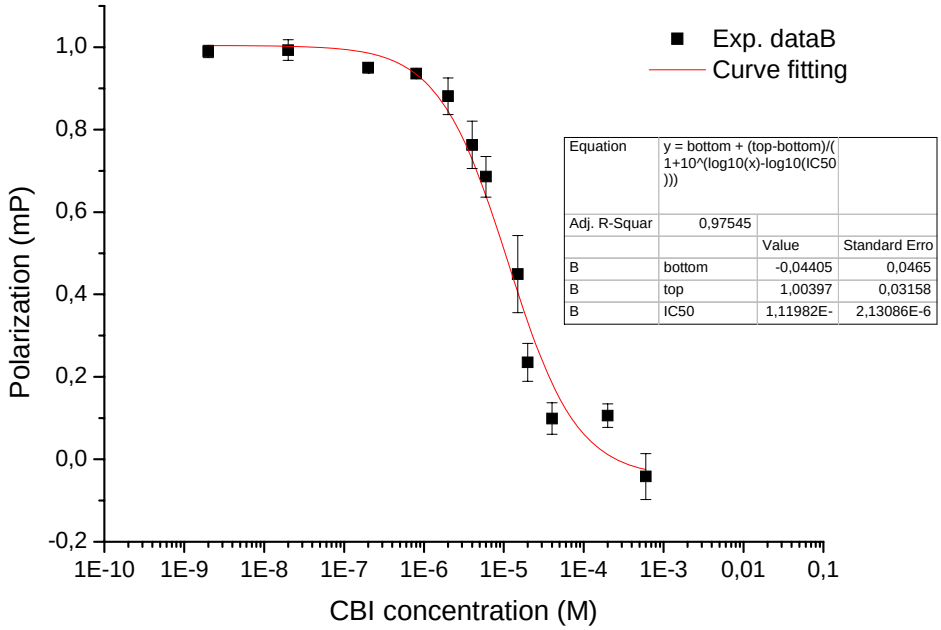

Figure S2. Curve of polarizarion obtained for the competitive fluorescence assay of Apa 1. Polarization values in mP are plotted against the $\log_{10}$ of the inhibitor concentration. The resulting sigmoid was fitted to the equation 2. The IC_{50} value is determined from the iterated parameters calculated for the best fitting of the curve to the experimental data.

The competitive binding of the miniproteins were measured 2 or 3 times in independent duplo experiments. The fluorescence polarization data of each miniprotein were used to calculate the IC_{50} using equation 2.

$$P = P_{\min} + \frac{\left(P_{\max} - P_{\min}\right)}{\left(1 + 10^{\left(\log(x) - \log(IC_{50})\right)}\right)} \qquad \text{(Eq. 2)}$$

P = polarization ; P_{min} = minimum value of polarization, P_{max} = maximum value of polarization

With the aid of equation 3, the K_d of the fluorescein labeled peptide and the IC_{50} values of each miniprotein, the K_i values for the different miniproteins could be determined.

$$K_i = \frac{IC_{50}}{1 + \frac{A_0 \cdot \left(y_0 + 2\right)}{2 \cdot K_d \cdot \left(y_0 + 1\right)} + y_0} - K_d \cdot \left(\frac{y_0}{y_0 + 2}\right) \qquad \text{(Eq. 3)}$$

IC_{50} : inhibitor concentration that reduces binding of the labeled peptide by 50%; *Kd* : dissociation constant of the protein- labeled peptide complex

$$y_0 = \frac{AB}{A}$$

AB is the concentration of protein-peptide complex and *A* the concentration of unbound peptide.

[1] E. Hur1, S.J. Pfaff, E.S. Payne, H. Grøn, B.M. Buehrer, R.J. Fletterick, *PLoS Biology*, 2004, **2**, e274.